Sequential collaboration network with sentiment coloring

Nielsen, Finn Årup

Publication date:
2013

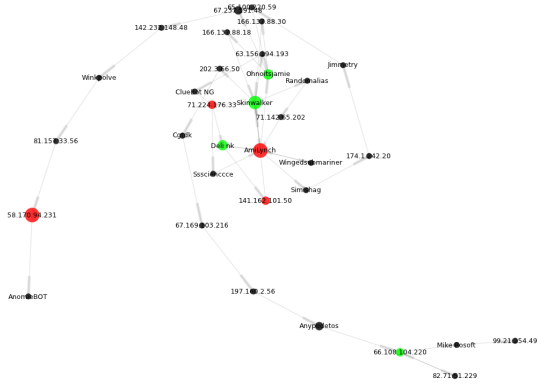
Document Version
Publisher's PDF, also known as Version of record

Sequential collaboration network with sentiment coloring

Finn Årup Nielsen, DTU Compute, Kongens Lyngby, Denmark

When multiple authors one after another edit a work and a revision control system records the revision history one may construct the so-called "sequential collaboration network" (SCN) [1], also called the "article trajectory" [2]. The nodes of an SCN each corresponds to one author and a directed link between two nodes indicates that an author has edited the work right after another author, thus the construction of an SCN results in a directed graph. As an example of a SCN I have implemented an online on-demand service for the revision history of individual Danish Wikipedia articles, presently available from http://toolserver.org/~fnielsen/cgi-bin/revvis. I let the node size correspond to the number of edits of the user has made.

In a research project we examine the positive and negative edits made on Wikipedia to a set of company articles on the English Wikipedia by different users. Continuously monitoring Wikipedia and adding the text data to our local database via the Wikipedia API, we use sentiment analysis [4] on the retrieved text to determine whether a user has added or removed positive or negative text from a Wikipedia article. The sentiment analysis method is presently based on a valence-labeled word list, AFINN, that we created ourselves [3].



How can we create a good overview over the sentiment analyzed data that both has a user and a temporal dimension? We produce plots of the sentiment evolution through time (edits) and barplots of summarized user sentiment towards a company as well as color coded lists of text "diffs". But to get a visualization of the mutual edits of users we render the data as color-coded SCNs where the node color is determined from summarized user sentiment.

As we for each company on Wikipedia monitor several articles the SCN for each company may have disconnected components if the users behind edits do not overlap between company articles. Visualizing the data we construct two plots: one with the entire SCN for the company and one "zoomed" plot with only the largest connected component.

We let the Python package NetworkX handle the graph layout and rendering. For connected SCNs we find that a two-stage layout procedure works well in many cases with small SCNs: We initialize the node position with a spectral layout in the first stage and then in the second stage apply a few iterations of force-based layout. For disconnected SCNs we only use the force-based layout.

The figure shows an automatically generated static plot with an SCN for a company. In the center is a "negative" user (red) directly connected to three "positive" users (green). An examinination of their edits reveals a small so-called "edit war" between the users, where the negative user adds text with negative sentiment towards the company on its product articles and where the positive users removes the inserted negative information. This is one example of the plots from several hundred companies we generate by batch and make available online at http://rb.compute.dtu.dk.

Acknowledgment: Responsible Business in the Blogosphere.

References

[1] T. Iba, K. Nemoto, B. Peters, and P. A. Gloor. Analyzing the creative editing behavior of Wikipedia editors: through dynamic social network analysis. *Procedia*, 2(4):6441–6456, 2010.

[2] B. Keegan, D. Gergle, and N. Contractor. Staying in the loop: structure and dynamics of Wikipedia's breaking news collaborations. In *WikiSym 2012*, 2012.

[3] F. Å. Nielsen. A new ANEW: evaluation of a word list for sentiment analysis in microblogs. In *Proceedings of the ESWC2011 Workshop on 'Making Sense of Microposts'*.

[4] B. Pang and L. Lee. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2):1–135, 2008.